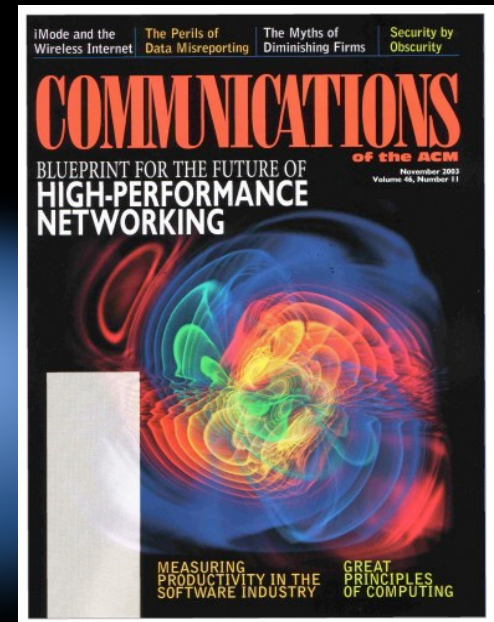


Transformation of Science through Cyberinfrastructure

Edward Seidel
Director, Office of Cyberinfrastructure
National Science Foundation
hseidel@nsf.gov



National Science Foundation
Where Discoveries Begin

Edward Seidel
hseidel@nsf.gov

Office of
Cyberinfrastructure

Data-Driven Multiscale Collaborations* for Complexity

Great Challenges of 21st Century

- Multiscale Collaborations
 - General Relativity, Particles, Geosciences, Bio, Social...
 - And all combinations...
- Science and Society being transformed by CI and Data
 - Completely new methodologies
 - “The End of Science” (as we know it)
- CI plays central role
 - No community can attack challenges
 - Technical, CS, social issues to solve
- *Places requirements on computing, software, networks, tools, etc*



The End of Science

The quest for knowledge used to begin with grand theories. Now it begins with massive amounts of data. Welcome to the Petabyte Age.



*Small groups still important!

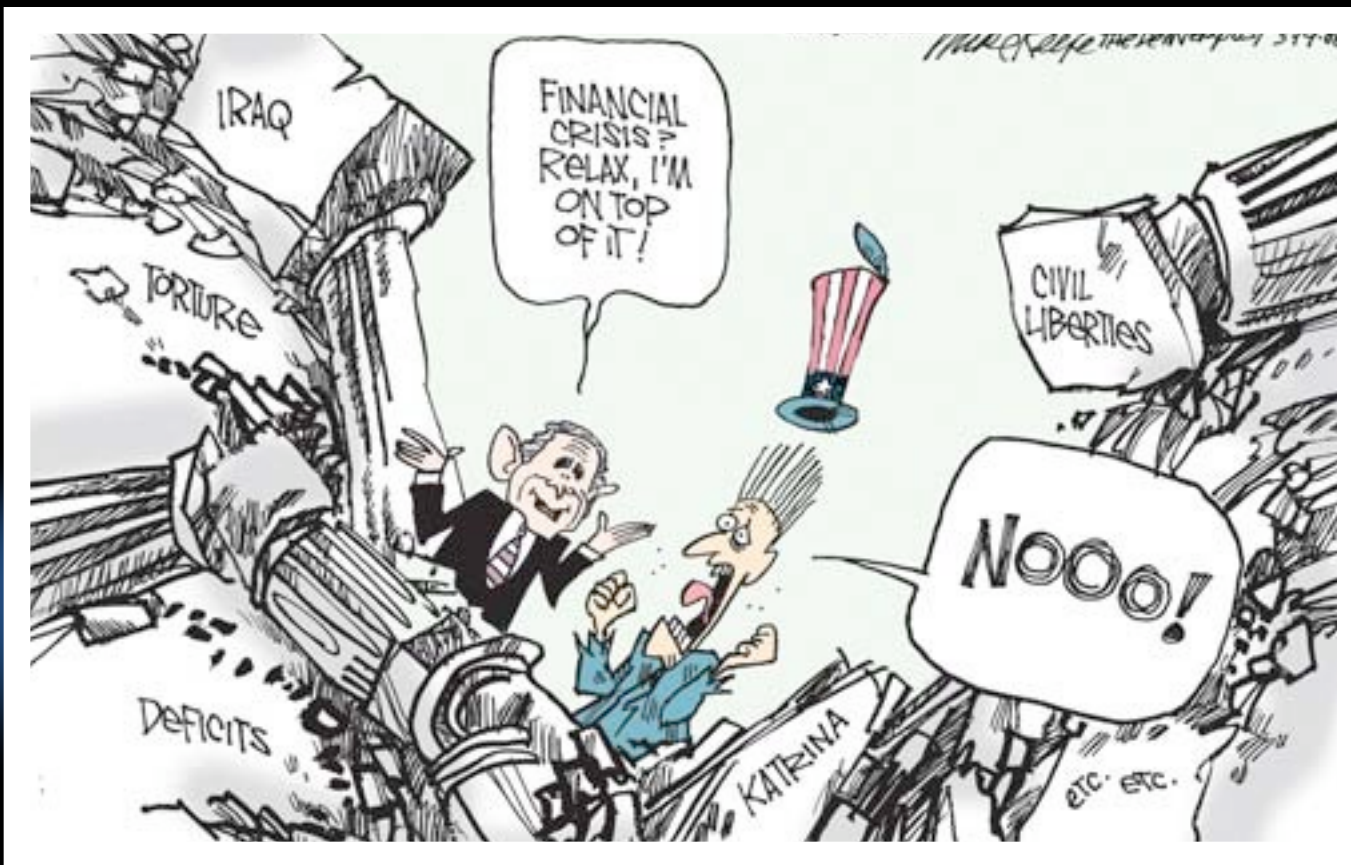


National Science Foundation
Where Discoveries Begin

Edward Seidel
hseidel@nsf.gov

Office of
Cyberinfrastructure

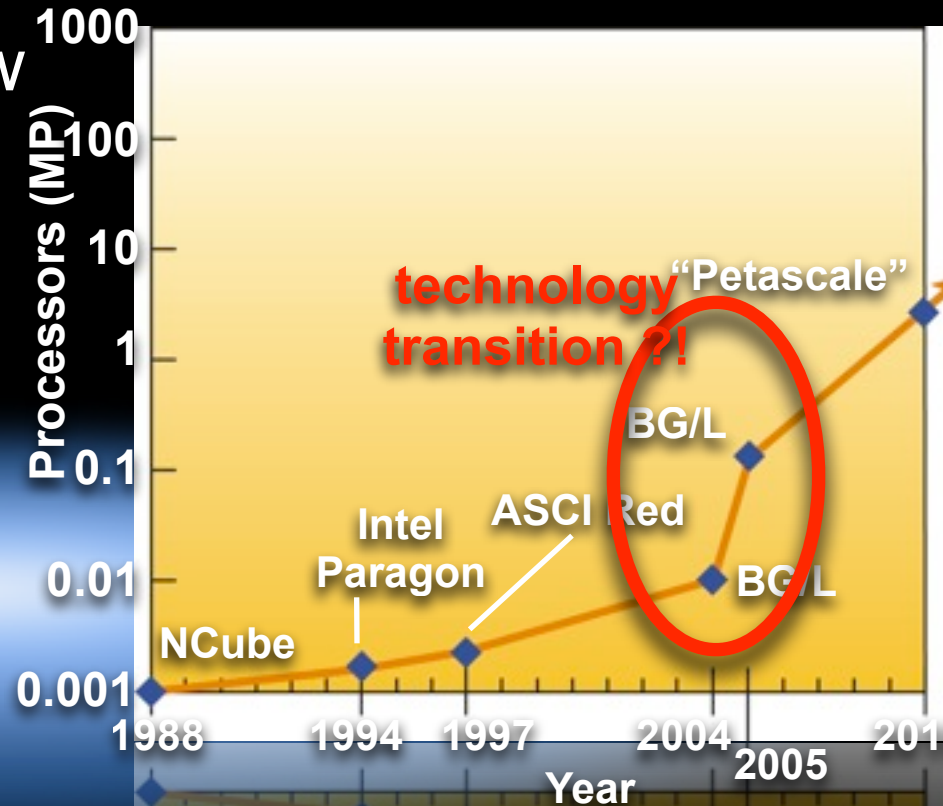
Crises to Deal With



Technology Crisis

(Adapted from Simon/Shalf)

- “The processor is the new transistor” (Patterson)
 - BG/L at LLNL: as many procs as transistors in the MC68000
 - N_{procs} has made a transition
 - Programming parallel codes like assembly language moving bits between transistors
- Multicore is coming on fast
 - N_{cores} doubles every 18 months, clock rate more or less fixed
 - Programming: MPI within a chip???, No: MPI + ???



Question: “How fault tolerant is your code?”

• Many CS, CI Challenges

National Science Foundation
Where Discoveries Begin

Edward Seidel
hseidel@nsf.gov

Office of
Cyberinfrastructure



Data Crisis: Information Big Bang

PCAST Digital Data

Industry

Wired, Nature

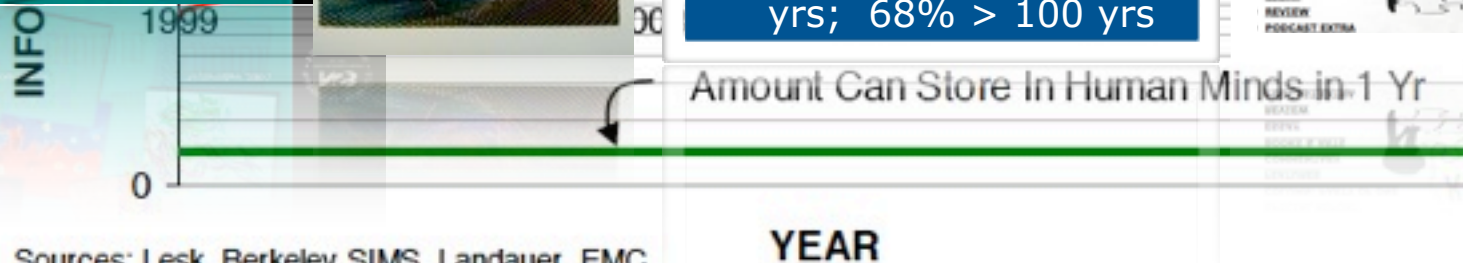


NSB Report: Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century



Storage Networking Industry Association (SNIA) 100 Year Archive Requirements Survey Report

"there is a pending crisis in archiving... we have to create long-term methods for preserving information, for making it available for analysis in the future." 80% respondents: >50 yrs; 68% > 100 yrs



Software Crisis

- Computers are exceedingly complex
 - Desktops with hundreds of cores
 - Supercomputers with millions of cores
 - They last 3-4 years...
- Software systems and applications
 - Science apps have 10^3 to 10^{6+} lines, have bugs
 - Applications may take decades to develop
 - We spend at least 10x as much on hardware
 - *GC communities place requirements on software for complex CI (not just HPC!)*
- We have a *crisis* in software
 - We don't know how to write it!
 - Is our science reproducible? If not...not science!



Toolkit for
complex CI?



NSF Vision

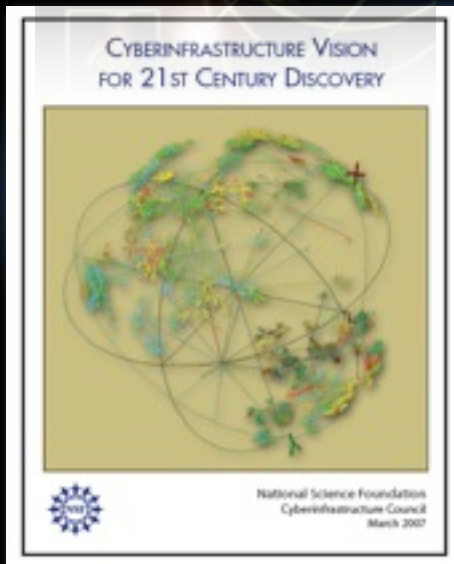
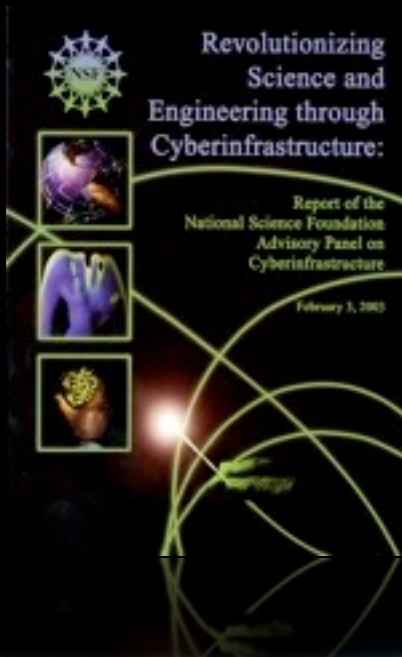
“National-level, integrated system of hardware, software, data resources & services... to enable new paradigms of science”

Virtual
Organizations for
Distributed Communities

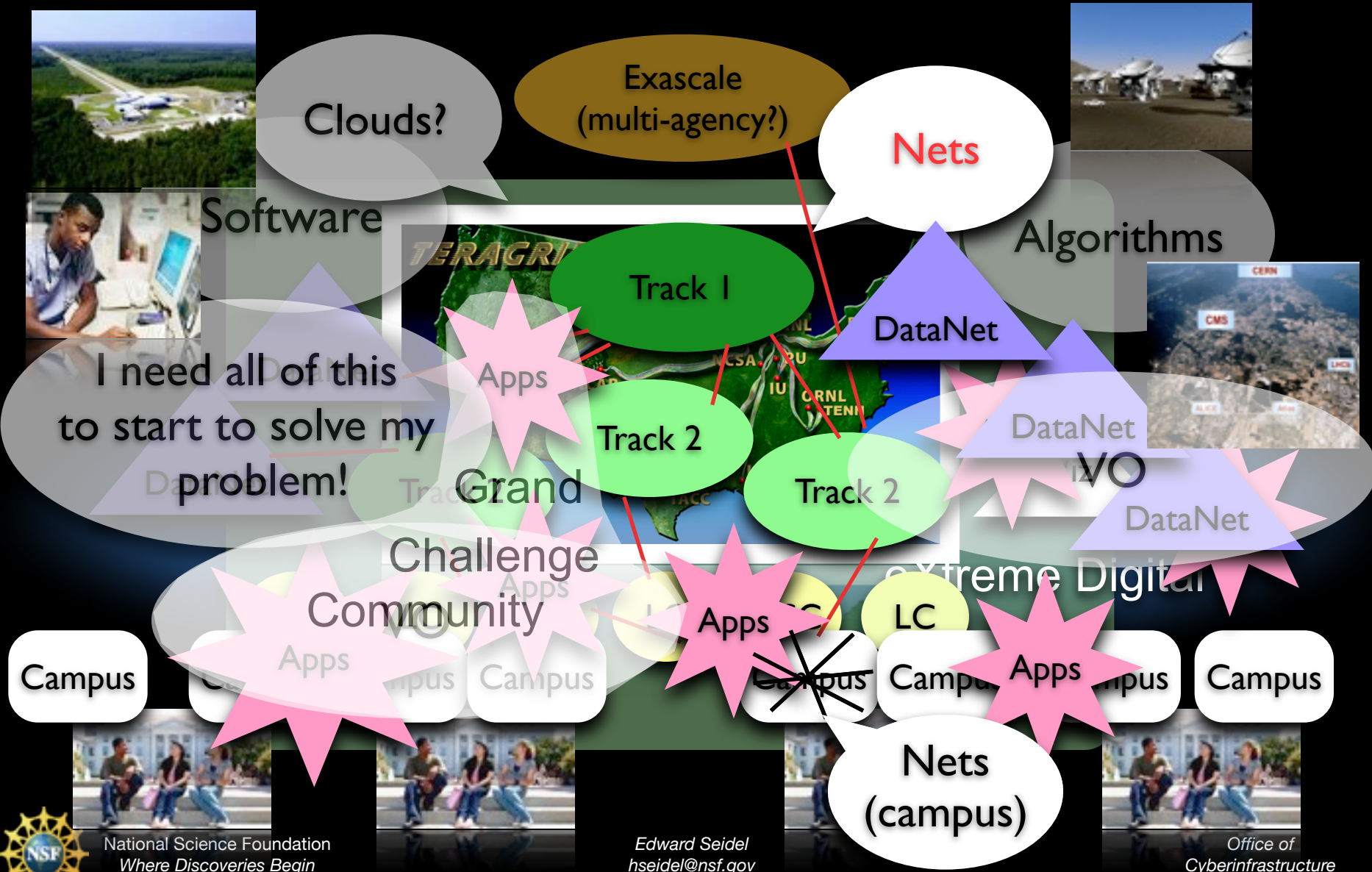
High
Performance
Computing

Data &
Visualization/
Interaction

Learning & Work Force
Needs & Opportunities



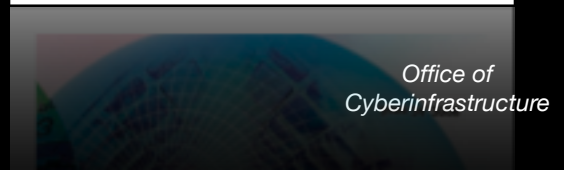
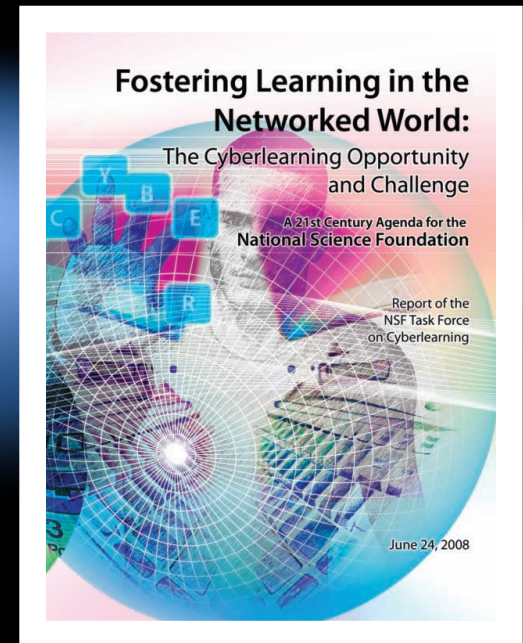
National CI Blueprint



Education Crisis

- The CI environment is running away from us!
- How do we develop a workforce to work effectively in this world?

We must be much more aggressive about training people to use this, to deal with this data, and to work in this collaborative environment



National Science Foundation
Where Discoveries Begin

What about Campuses?

- Next generation science requires *huge* data to be manipulated, retrieved, stored, analyzed
- Collaborative environments will need unprecedented levels of sophistication
- Can barely do low-end video conferencing today! HD, Optiportal-level environments needed
- *We need to seriously rethink our campus environments and how they can support new modalities of research, collaboration, education*
- Federated ID management!



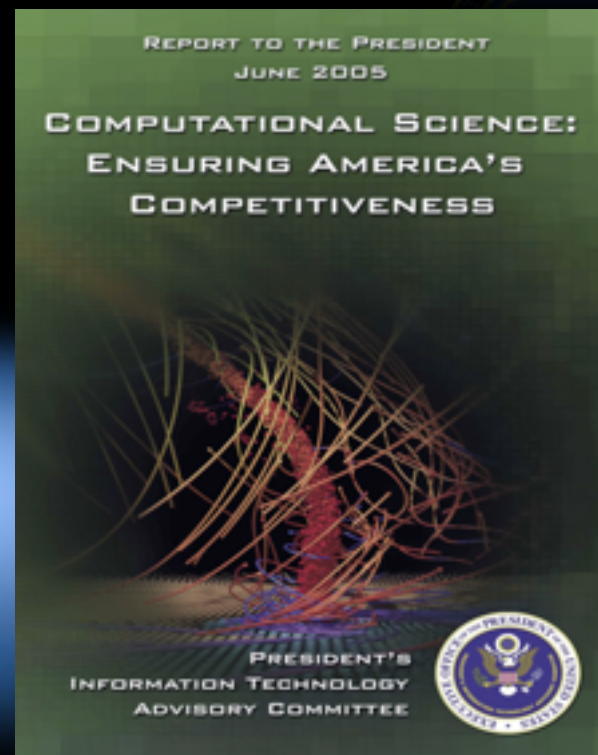
National Science Foundation
Where Discoveries Begin

Edward Seidel
hseidel@nsf.gov

Computational Science PITAC Report Summary



*“Together with theory
and experimentation,
computational science
now constitutes the
“third pillar” of scientific
inquiry...”*



National Science Foundation
Where Discoveries Begin

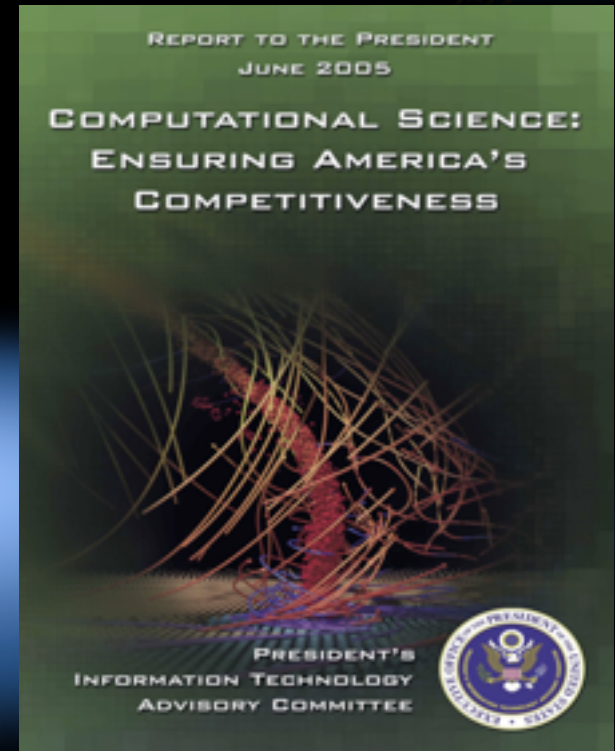
Edward Seidel & Fahmida Chowdhury
hseidel@nsf.gov & fchowdhu@nsf.gov

Office of Cyberinfrastructure
Social, Behavioral & Economic
Sciences Directorate

Computational Science PITAC Report Summary



“Universities must significantly change organizational structures: multidisciplinary & collaborative research are needed [for US] to remain competitive in global science”



National Science Foundation
Where Discoveries Begin

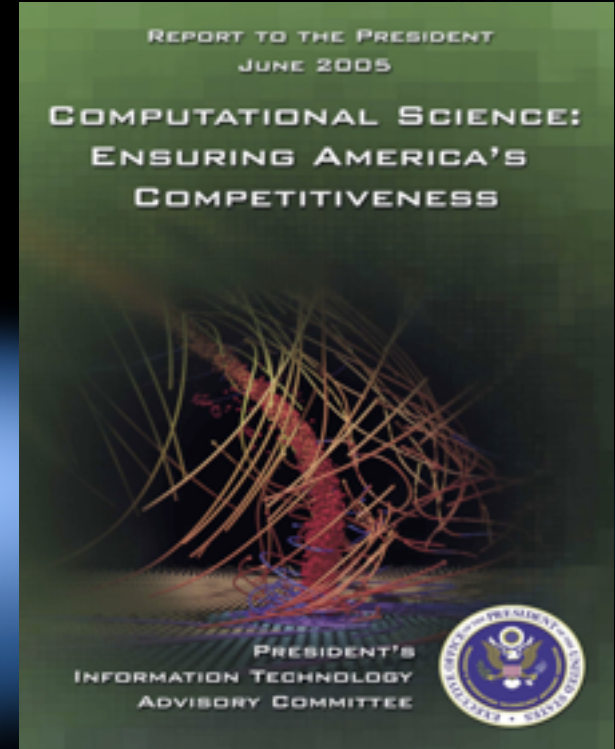
Edward Seidel & Fahmida Chowdhury
hseidel@nsf.gov & fchowdhu@nsf.gov

Office of Cyberinfrastructure
Social, Behavioral & Economic
Sciences Directorate

Computational Science PITAC Report Summary



“Itself a discipline, computational science serves to advance all of science.....inadequate and outmoded structures within the Federal government and the academy today do not effectively support this critical multidisciplinary field”



National Science Foundation
Where Discoveries Begin

Edward Seidel & Fahmida Chowdhury
hseidel@nsf.gov & fchowdhu@nsf.gov

Office of Cyberinfrastructure
Social, Behavioral & Economic
Sciences Directorate

Rough Plan for OCI

- Attempt to fill out holes in current CI coverage
 - Data, software, algorithms, networks...
 - HPC and beyond: Clouds, campus integration...
- Workforce development!
 - Many exponentials: data, compute, collaboration
- Task Forces: Campus Bridging, Education, Computing, Software, Data, GC Communities
- OCI as base for CI development, deployment & *computational science* research and education
- New programs: GC communities attack complex problems at frontiers of science and engineering
 - CI-driven, with above pieces in place first

